IOPscience iopscience.iop.org

Home Search Collections Journals About Contact us My IOPscience

Tunnelling in multidimensional systems

This article has been downloaded from IOPscience. Please scroll down to see the full text article.

1994 J. Phys. A: Math. Gen. 27 4969

(http://iopscience.iop.org/0305-4470/27/14/023)

View the table of contents for this issue, or go to the journal homepage for more

Download details:
IP Address: 171.66.16.68
The article was downloaded on 01/06/2010 at 21:31

Please note that terms and conditions apply.

J. Phys. A: Math. Gen. 27 (1994) 4969–4993. Printed in the UK

Tunnelling in multidimensional systems

Stephen C Creagh
Niels Bohr Institute, Blegdamsvej 17, DK-2100, Copenhagen, Denmark

Received 8 March 1994, in final form 28 April 1994

Abstract. Tunnelling is examined from the point of view of analytic continuation of Lagrangian manifolds. This enables an intrinsically multidimensional approach to many problems. We illustrate the ideas by computing energy splittings between congruent tori in a number of integrable systems. The examples include an integrable but non-separable system which cannot be solved by reduction to one-dimensional systems. The calculations can all be viewed in a geometrical way that does not make explicit use of separation of coordinates.

1. Introduction

While tunnelling is very naturally incorporated into WKB analyses of one-dimensional systems [1, 2], the understanding of tunnelling processes in many dimensions (see Gutzwiller [3] and references therein) is much less complete. Part of the difficulty of generalizing one-dimensional calculations to higher dimensions is that the one-dimensional theory is usually presented in such a way that it relies heavily on the fact that one is typically interested in finding eigensolutions of Hamiltonians of the kinetic-plus-potential type—so the Schrödinger equation is a second-order linear ordinary differential equation. We adopt a point of view here that avoids such explicit assumptions about the Schrödinger equations and as a result it is much easier to generalize to multidimensional calculations.

The essential point is that WKB approximations always arise from the identification of a quantum object—a wavefunction, propagator or Green function for example—with a Lagrangian manifold in phase space [3]. By focusing attention on the Lagrangian manifold rather than the Schrödinger equation itself, one can calculate tunnelling effects in a unified way that is not qualitatively different for Schrödinger equations of different order, or even different dimension. In simple WKB calculations, the Lagrangian manifold is an n-dimensional submanifold of $2n$-dimensional phase space—tunnelling corrections are then included by complexifying the Lagrangian manifold and phase space, and allowing complex phase-space points into the reconstruction of the quantum object.

The multidimensional problems that have thus far been solved by WKB techniques [3] are: (i) quantization of integrable systems, (ii) scattering and (iii) approximation of propagators and Green's functions, and subsequent calculation of trace formulae. In this paper we concentrate on (i) as a source of examples. In integrable systems the eigenstates are constructed from invariant tori—quantization conditions arise from demanding that the wavefunction remain single-valued after encircling all closed loops on the manifold. To quantize the system with tunnelling corrections we regard the first integrals as functions of complex variables and construct the complexified manifold as a level set of these complex functions—the complexified manifold has $2n$ real dimensions whereas the real manifold is n-dimensional. One then similarly demands single-valuedness of the wavefunction after

0305-4470/94/144969+25$19.50 © 1994 IOP Publishing Ltd

traversing all closed loops on this higher-dimensional manifold. Viewing the procedure in this way has the advantage that one comes closer to treating the problem in a canonically invariant way. The complex Lagrangian manifold and its homological structure are geometrical invariants.

We review the one-dimensional theory in section 2 and present it in such a way that the dependence on dimension is minimized. We then give the generalization to many dimensions in section 3. The theory is not substantively different from the one-dimensional theory, but is presented in such a way that intrinsically multidimensional calculations are possible.

In real WKB calculations the changes that occur in the wavefunction over a closed loop can be expressed in a canonically invariant way. The same must also be true of the complex calculation but this has not been shown explicitly. There is a step missing which corresponds roughly to the proof by Arnold [4] in the real case that the Maslov index for a closed loop is a canonical invariant. It is possible, however, to calculate the wavefunction changes in a specific set of coordinates and we do this in this paper for numerous examples. We present some simple rules that are valid under the special circumstance that all of the important structure of the complex Lagrangian manifold can be understood by taking a slice through it along real configuration space. All of the examples presented in this paper fall into this class. The rules amount to a correction of the usual Maslov shift that depends on the extent to which exponentially growing solutions are present on the forbidden side of a caustic.

The simplest case where tunnelling plays a role in integrable systems is that there are two symmetric tori confined to distinct wells in a potential. A two-fold degeneracy in EBK levels is then broken when tunnelling effects are included. We present a series of examples of this sort and in each case the essential structure is of a common type—when one examines a real-position slice through the complex manifold one finds that the two real tori are connected by one or more complex tori in the forbidden region. Single-valuedness of the wavefunction around all closed loops on the real and complex tori yields the energy splitting. Note that this picture is independent of whether or not the system is separable and so we can solve the problem in an intrinsically multidimensional way, without resorting to reduction to one-dimensional systems. We illustrate this by solving a particular integrable, but non-separable, problem. The calculation for this system is qualitatively the same as for separable systems—and in each case the energy splitting scales with $\hbar$ as $\hbar \exp(-K/\hbar)$, where iK is a complex action. The tunnelling results here apply to a different regime to those of Wilkinson [5, 6] where the continuations of the real tori are assumed to meet along a surface of just one complex dimension rather than matching smoothly to the same complex manifold—there the splittings are calculated to scale as $\hbar^{3/2} \exp(-K/\hbar)$.

Calculations of splittings for non-separable systems have also been discussed by Meyer *et al* [7, 8], where emphasis is placed on the relationship of the approximation to complex trajectories. By contrast, here the emphasis is on the global structure of the associated Lagrangian manifolds. We believe that this approach is preferable for an intrinsic and (eventually) coordinate-free solution of the problem.

Let us summarize the paper. In section 2 we review one-dimensional tunnelling theory, partly to establish notation, and partly to set up for generalization to the multidimensional case. In section 3 we discuss the multidimensional case and give the correction to the Maslov shift that accounts for tunnelling. We proceed to illustrate the theory with a number of integrable systems: in section 4 we start by solving the one-dimensional double well, in section 5 a separable two-dimensional well, in section 6 a non-separable two-dimensional well, and in section 7 the problem of H_2^+, which is separable. Section 6 also contains a brief overview of the solution of integrable problems in general.

2. Review of tunnelling in one dimension

The semiclassical theory of tunnelling is very well understood for one-dimensional problems of the kinetic-plus-potential type. Since the theory we will present for multidimensional systems will be a simple generalization of the one-dimensional method, it is useful to begin by briefly outlining this one-dimensional theory. See Berry and Mount [1] for a good elementary discussion and for further references. Here we are interested only in the lowest-order expansions in $\hbar$ for each solution of the Schrödinger equation, but complete expansions have also been investigated for quartic oscillators [9, 10].

We begin by writing the WKB approximation for a simple one-dimensional Hamiltonian of the type

$$H = p^2/2 + V(q) \tag{1}$$

where we might typically take $V(q)$ to be a double well potential. Now, the Schrödinger equation for this Hamiltonian in position representation is a second-order linear ODE and has two independent solutions. The WKB approximation for these solutions is obtained by solving for the two branches of momentum as a function of position,

$$p_\pm(q) = \pm\sqrt{2(E - V(q))} \tag{2}$$

integrating to get the action,

$$S_\pm(q) = \int_{q_0}^{q} p_\pm(q')\,\mathrm{d}q' \tag{3}$$

and constructing the two approximate solutions as follows:

$$\psi_\pm(q) = p_\pm(q)^{-1/2} \exp\left[\frac{\mathrm{i}}{\hbar} S_\pm(q)\right]. \tag{4}$$

A general solution to the differential equation is locally expressed as a linear combination of these two solutions,

$$\psi(q) = a_+\psi_+(q) + a_-\psi_-(q) \tag{5}$$

where a_+ and a_- are constants determined by the boundary conditions.

The solution of (5) is valid except at turning points, where $p_\pm(q) = 0$, the solutions $\psi_\pm(q)$ diverge and the WKB approximation is invalid. Away from these regions, it might be expected that the coefficients $a_\pm$ should remain constant, since this is what is required to satisfy the Schrödinger equation. This is not entirely true however. If we regard q as a complex variable and extend the solution (5) to the complex plane, it turns out it is necessary to allow the coefficients to change discontinuously along lines in the complex plane (of real codimension 1) in order that a globally single-valued solution be obtained. These lines are called Stokes lines and emanate from the turning points. Though these lines are intimately connected with the turning points, it is not necessary to actually be close to a turning point to cross a Stokes line and have the coefficients change. This inability to predict changes in $a_\pm$ purely from local conditions is the main source of difficulty in calculating tunnelling effects in an elegant way.

A complete explanation of the reasoning behind the existence of Stokes lines and a detailed explanation of their structure can be found in [1, 2]. Here we will briefly outline their structure and computation. To understand the Stokes lines, it is necessary to examine the behaviour of a solution in the neighbourhood of a turning point. Let us choose q_0 in (3) to be the (possibly complex) turning point. The actions $S_\pm(q)$ will typically have non-zero imaginary parts and therefore when they are substituted into (4) one of

the fundamental solutions will be exponentially larger than the other (the ratio is of order $\exp(-|\Im(S_+ - S_-)|/\hbar)$). The smaller solution is called subdominant and the larger dominant. The Stokes lines are defined as the set where the dominant solution is maximally dominant, i.e. where $S_+ - S_-$ is pure imaginary. Through a local expansion of $p(q)$ around the turning point it is possible to see that the Stokes lines consist of three lines emerging from the turning point at angles of 120°.

On a Stokes line, a change in the coefficient of the subdominant solution will have the least disruptive effect on the overall solution (5) because it is there that the subdominant solution is smallest in a relative sense. It is found that in order for the solution to be single-valued around the turning point, it is necessary, when passing through a Stokes line, to change the subdominant coefficient by an amount proportional to the dominant coefficient as follows:

$$a_\alpha \to a_\alpha + c a_\beta \tag{6}$$

where $\psi_\alpha(q)$ is subdominant to $\psi_\beta(q)$. The coefficient c is referred to as the Stokes constant, and it's value depends on phase conventions for the amplitudes $p_\pm(q)^{-1/2}$, the direction of crossing of the Stokes line etc. In the appendix we outline a set of phase conventions for which $c = 1$ always. Other standard phase conventions [1] give $c = \pm i$.

A complete solution of the problem, including tunnelling effects, is obtained by globally matching the coefficients $a_\pm$ over the complex plane and choosing the energy and coefficients in such a way that $\psi(q) \to 0$ as $|q| \to \infty$ along the real axis.

This is an overview of the most common case of tunnelling in one dimension and complete discussion of the solution of typical problems can be found in, for example, Berry and Mount [1]. It is useful now to generalize the discussion to one-dimensional problems where the Hamiltonian can be an arbitrary function of q and p, not necessarily of the kinetic-plus-potential form shown in (1). To do this, and to generalize further to many-dimensional problems, it is good to view the process in a way that minimizes the dependence on a specific choice of representation.

We see that there is one local independent solution $\psi_\alpha(q)$ of the Schrödinger equation for each branch $p_\alpha(q)$ of the implicit equation $H(q, p) = E$. An appropriate generalization of (4) to arbitrary Hamiltonians is

$$\psi_\alpha(q) = \left(\frac{\partial H}{\partial p}\right)_\alpha^{-1/2} \exp\left[\frac{i}{\hbar} S_\alpha(q)\right]. \tag{7}$$

So we can associate one independent solution with each branch of the Riemann sheet structure defined by $H(q, p) = E$. Therefore the solutions can be regarded as defining a single-valued function on the Riemann sheet manifold, albeit with discontinuous changes at Stokes lines (and along branch cuts of the amplitude, as discussed in the appendix). Note that the Riemann sheet manifold defined by $H(q, p) = E$ can be regarded as an intrinsic subset of complexified phase space and is independent of choice of coordinates. The quantization conditions come from traversing all topologically distinct loops on this manifold and demanding that the solution return to its original value at the end of each loop. This statement is also independent of representation, since the topology of the Riemann sheet is independent of which coordinates the manifold is projected onto.

We illustrate this invariant structure in figure 1 for the case of a quartic potential $V(q) = -q^2/2 + q^4/4$, where E lies below the local maximum. There are four turning points and we choose the branch cuts of $\sqrt{(2(E - V))}$ to run between pairs of turning points as in figure 1(a). By identifying the sheets along the branch cuts we get the global structure shown in figure 1(b). While the intermediate construction of the two Riemann

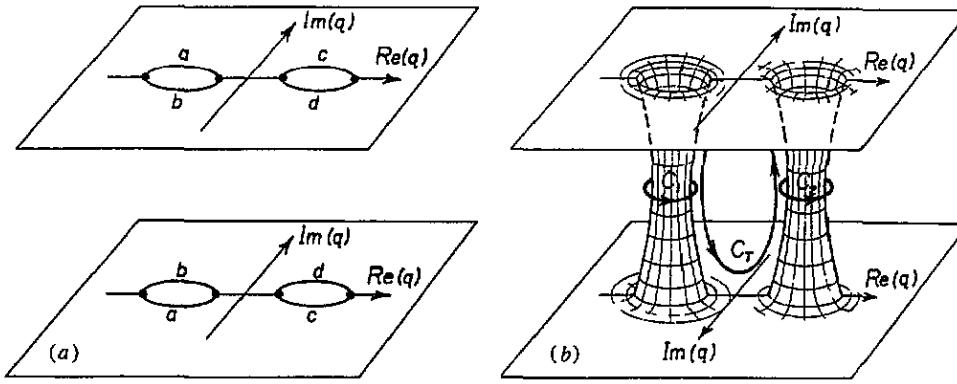

Figure 1. Here is the Riemann sheet for the double well potential. In figure 1(a) we show two separate sheets with branch cuts connecting pairs of turning points. After gluing the sheet along the identified edges (a, b, c, d) one gets the structure in figure 1(b) (it is necessary to turn one of the sheets upside down to glue along the proper edges). We show the three homologically distinct loops C_1, C_2 and C_T of the manifold in figure 1(b).

sheets in figure 1(a) is particular to the q-representation, the structure shown in figure 1(b) is obtained no matter which representation is used. There are three independent loops, C_1, C_2 and C_T, also shown in figure 1(b). Demanding single-valuedness after looping around C_1 and C_2 gives approximate EBK quantization of the two wells, and single-valuedness around C_T completes the computation of tunnelling corrections.

The representation dependence of the calculation comes from the details of actually computing changes to the solutions while traversing the loops. Changes in the solutions come from crossing Stokes lines—their positions, as well as the way the wavefunction changes at them, depends intimately on the representation. The situation is analogous to the quantization of tori in multidimensional integrable systems, where the position and nature of each Maslov index shift depends on the representation, even though the total shift at the end of a closed loop is representation-independent [4, 11]. Similarly, we expect that in the present case the total change in the coefficient of a fundamental solution over a closed loop can be expressed in a representation-independent way. Unfortunately, the non-local nature of the changes in this case makes it much much more difficult to analyse than in the real case.

Turning points in the general case are points at which branches of $p(q)$ coalesce—not necessarily at $p(q) = 0$—and the generic case is that just two branches, $p_\alpha(q)$ and $p_\beta(q)$ say, coalesce. Stokes lines radiate outwards along the relevent two branches, defined by the condition that $S_\alpha - S_\beta$ be imaginary (the integration constant q_0 in the analogue of (3) is taken to be the turning point). On crossing a Stokes line the subdominant solution, $p_\alpha(q)$ say, changes as before according to (6).

There is a fundamental difference between the general case and the kinetic-plus-potential case, however. As pointed out by Berk *et al* [12], Stokes lines can originate not only at turning points, but also at the intersection of Stokes lines that connect different pairs of branches (this cannot happen for (1) because there is only one pair of branches). Suppose ψ_α, ψ_β and ψ_γ represent three distinct branches and that two Stokes lines, $\mathcal{S}_{\alpha\beta}$ connecting the (α, β) branches, and $\mathcal{S}_{\beta\gamma}$ connecting the (β, γ) branches, intersect at q_0. Suppose further that ψ_α is dominant over ψ_β and ψ_β is dominant over ψ_γ on their respective Stokes lines. Using just the rules outlined so far, one gets different results after passing through both Stokes lines, depending on the order in which they are crossed. Suppose one starts

with solution ψ_α, passes through first $\mathcal{S}_{\alpha\beta}$, and then $\mathcal{S}_{\beta\gamma}$—then the solution emerges as $\psi_\alpha + c_{\alpha\beta}\psi_\beta + c_{\alpha\beta}c_{\beta\gamma}\psi_\gamma$, where $c_{\alpha\beta}$ and $c_{\beta\gamma}$ are Stokes constants. If, on the other hand, one crosses $\mathcal{S}_{\beta\gamma}$ first and then $\mathcal{S}_{\alpha\beta}$, the solution emerges as $\psi_\alpha + c_{\alpha\beta}\psi_\beta$. The resolution of this inconsistency lies in the fact that one must insert another Stokes line connecting the (α, γ) branches, starting at q_0 and bisecting the sector in which $\mathcal{S}_{\beta\gamma}$ is crossed before $\mathcal{S}_{\alpha\beta}$. The new Stokes line makes up for the missing term in that case. Full details of the calculation of these Stokes lines and the corresponding Stokes constants are given in [12]. In this paper we deal with examples for which this kind of Stokes line is not important, but in general problems these Stokes lines will be needed for a complete analysis.

So this represents the structure of tunnelling calculations in one dimension. We are now ready to tackle higher-dimensional cases.

3. Generalization to many dimensions

In this section we examine an obvious generalization of the discussion in one dimension to more dimensions.

Let us start by assuming that some state in a multidimensional problem is determined by a Lagrangian manifold λ. If we were to ignore tunnelling effects, then λ would be taken to be a real n-dimensional submanifold of real $2n$-dimensional phase space. In the example of the previous section this real Lagrangian manifold would have been the real level set $H(q, p) = E$. The wavefunction is reconstructed by summing as follows over local solutions $\psi_\alpha(q)$, determined by branches α of the manifold that correspond to points in real phase space,

$$\psi_\alpha(q) = [\rho_\alpha(q)]^{1/2} \exp\left[\frac{\mathrm{i}}{\hbar} S_\alpha(q)\right]. \tag{8}$$

Here $\rho_\alpha(q)$ is a probability density on λ, determined by the particular problem, and $S_\alpha(q)$ is an action obtained by integrating the momentum over λ in an obvious generalization of (3). The Lagrangian condition states that the action should not change under continuous deformation of integration path (for more detail see [13]). In many dimensions the WKB approximation breaks down at caustics where two branches of the manifold coalesce and $\det(\partial p_\alpha/\partial q)^{-1} = 0$. These local solutions are patched together to form a global solution by including discrete phase shifts at caustics determined by the appropriate Maslov indices.

We now point out that it is not difficult to combine this picture with the one-dimensional techniques of the previous section to give a procedure for calculating tunnelling effects in many dimensions—in cases where the quantum state is known to be determined by a Lagrangian manifold (not all states have an obvious Lagrangian manifold associated with them—for example, eigenstates of chaotic systems). The solutions $\psi_\alpha(q)$ are analytically continued to complex q, as in the previous section. There are obvious analogues of the Stokes lines (in this case Stokes surfaces), on which the coefficents multiplying the local solutions change discontinuously and global matching of these coefficients, along with the demand that the solution decay at infinity, gives quantization conditions as before. The problem of course is that in practice the topology of these higher-dimensional structures will be much more difficult to analyse, but in principle this can be achieved.

As before it is desirable to view the procedure in as invariant a manner as possible. Rather than analytically continuing the solutions directly, let us think of continuing λ and the density ρ, and then reconstructing the continued wavefunction from these complexified objects. Analytical continuation of λ gives a surface of $2n$ real dimensions embedded in a complex phase space of $4n$ real dimensions—and a combination of the Lagrangian

property and Cauchy's theorem guarantees that integrals of $\boldsymbol{p} \cdot \mathrm{d}\boldsymbol{q}$ along λ are invariant under continuous deformation of path. Continuation of λ is easy if the manifold is known to be the level surface of a set of n analytic functions $F_\mu(\boldsymbol{q}, \boldsymbol{p})$, as is the case for integrable systems—one merely extends the functions to depend on complex variables $\boldsymbol{q}$ and $\boldsymbol{p}$ and constructs the level sets in complex phase space. Analytic continuation of the wavefunction is achieved simply by regarding the components of $\boldsymbol{q}$ as complex variables in (3), with one local solution being associated with each branch α of the complexified λ over $\boldsymbol{q}$.

To analyse the structure around a caustic and determine Stokes constants etc, we use normal coordinates to reduce the problem locally to a one-dimensional one. Suppose C is a caustic surface, obtained when λ is projected onto configuration space in some neighbourhood. We choose configuration space coordinates $(q_1, \ldots, q_n)$ such that C is specified locally by $q_1 = 0$. In fact it is possible to choose coordinates $(\boldsymbol{q}, \boldsymbol{p})$ on phase space in such a way that λ is specified by the conditions

$$\begin{aligned} p_1^2 &= -q_1 \\ p_i &= 0 \qquad i = 2, \ldots, n \end{aligned} \tag{9}$$

as, for example, in appendix 12 of Arnold [13]. In these coordinates the wavefunction decouples into a product

$$\psi_\alpha(q_1, q_2, \ldots, q_n) = \chi_\alpha(q_1)\phi_\alpha(q_2, \ldots, q_n) \tag{10}$$

where $\chi_\alpha(q_1)$ has a turning point at $q_1 = 0$ and $\phi_\alpha(q_2, \ldots, q_n)$ has no singular structure. Changes in coefficients of the solutions are then determined entirely by the behaviour of the one-dimensional functions $\chi_\alpha(q_1)$ and are independent of the coordinates $(q_2, \ldots, q_n)$. The coefficients of the solutions change when q_1 is on a Stokes line of $\chi_\alpha(q_1)$ and $(q_2, \ldots, q_n)$ are arbitrary—and the changes are determined from $\chi_\alpha(q_1)$ as outlined in the previous section.

From here we can reformulate the conditions for Stokes surfaces in a way that they are independent of the construction of normal coordinates. The Stokes surfaces have $2n-1$ real dimensions and are of real codimension 1 on the $2n$-dimensional complexified manifold λ. These Stokes surfaces radiate from caustic surfaces that are determined by the condition $\det(\partial p_\alpha/\partial q)^{-1} = 0$—caustic surfaces are of real codimension 2 on λ. If the Stokes surface corresponds to a caustic where two branches α and β coalesce, then it is determined by the condition that

$$\mathrm{Re}\left[S_\alpha(\boldsymbol{q}) - S_\beta(\boldsymbol{q})\right] = 0 \tag{11}$$

where the actions are obtained by integrating to $\boldsymbol{q}$ from a point on the caustic. In normal coordinates this happens to coincide with the conditions that q_1 lie on a Stokes line of $\chi_\alpha(q_1)$ and that $(q_2, \ldots, q_n)$ are arbitrary. The coefficients of the local solutions now change, as usual, according to (6). The Stokes constants could be determined by calculating in normal coordinates and using one-dimensional procedures. One could also calculate them from first principles by encircling a caustic in complex configuration space and demanding that the overall wavefunction be single-valued. Full details are presented in the appendix.

The preceding discussion deals only with the case of Stokes surfaces emanating from simple caustics where just one eigenvalue of the matrix $(\partial p_\alpha/\partial q)^{-1}$ passes through zero. In a generic calculation it will be possible to circumvent places where more than one eigenvalue passes through zero, so this restriction is not a real problem. See [14], however, for a discussion of Stokes surface structure in the neighbourhood of a degenerate point. More serious is the fact that it does not cover the case of Stokes surfaces arising from the intersection of two other Stokes surfaces as in the discussion at the end of the previous section. While in principle this case could also be dealt with by simple generalization of the one-dimensional discussion we will not do so here.

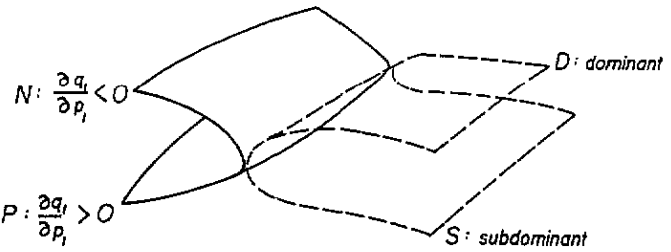

Figure 2. We show schematically a real-q slice through λ near a caustic. The branches on the real side are shown in full curves and are labelled N and P according to whether the derivative is negative or positive. The complex branches on the forbidden side are shown by broken curves and are labelled D and S for dominant and subdominant. The associated amplitudes are listed in (12)–(15).

To keep things simple, we will restrict ourselves to problems where the only important Stokes surfaces are those emanating from caustic surfaces that contain real configuration space points. In this case it is necessary only to analyse the wavefunction in a small neighbourhood of real configuration space. The essential structure of λ—at least the structure that is important for the determination of $\psi(q)$ for real q—can then be understood by taking a slice through λ corresponding to real q. This fact simplifies significantly the task of determining the topology of the system, and this is important because in many dimensions it becomes rapidly very difficult to get a complete understanding of the structure.

We will now write down simple rules for determining the relative amplitudes, phase factors etc, of the local solutions in a neighbourhood of a simple fold caustic at real values of q. These rules will amount to a simple modification of the usual Maslov phase shift that depends on the degree to which the non-decaying solution is present on the forbidden side of the caustic. We show schematically a slice through λ at real q in figure 2 for the case of two degrees of freedom. On one side of the caustic there are two real solutions for each q, represented in figure 2 by full curves—we label these branches P and N according to whether $\partial q_1/\partial p_1 > 0$ or $\partial q_1/\partial p_1 < 0$, respectively. These derivatives are understood to be taken while $(q_2, \ldots, q_n)$ are held fixed. On the forbidden side of the caustic there are two complex solutions represented by broken curves in figure 2—we label these by D and S, respectively, according to whether the appropriate branch is dominant or subdominant.

We will next specify the relative amplitudes a_α of each of the fundamental solutions $\psi_\alpha(q)$. First it is necessary to specify in detail the phase conventions in (8). We choose the symbol $[\rho_\alpha(q)]^{-1/2}$ to represent the analytic function that coincides with $|\rho_\alpha(q)|^{-1/2}$ when q is real. Of course, as one moves around the caustic in complex configuration space, this identification becomes ill defined as the branches exchange identity—but is well defined sufficiently close to real configuration space. Next, we will suppose that the path of the integral defining $S_\alpha(q)$ starts on a given point in λ—so $S_\alpha(q)$ defines a function that is single-valued on λ in a neighbourhood of the caustic and becomes multi-valued only after projection to configuration space. With these conventions, the amplitudes are

$$a_S = 1 \tag{12}$$

$$a_D = T \tag{13}$$

$$a_N = e^{i\pi/4}\left(1 - \tfrac{1}{2}iT\right) \tag{14}$$

$$a_P = e^{-i\pi/4}\left(1 + \tfrac{1}{2}iT\right). \tag{15}$$

These equations are derived in the appendix using a simple extension of one-dimensional

arguments. We have the freedom to choose a common normalization factor, and this has been done in such a way that $a_S = 1$. Once this is done, then a_D is left undetermined by a purely local analysis and we equate this to the temporarily unknown variable T, which we will refer to as the tunnelling parameter. T must later be determined by a global matching of the solutions over the whole of λ. Once these two choices have been made the remaining two coefficients a_N and a_P are determined completely, as in (14) and (15).

These amplitudes are valid for real q. Real configuration space is contained in a Stokes surface on the forbidden side, so a_S changes discontinuously when q becomes complex. If one moves off real configuration space in the direction where one ends up on the N branch after encircling the caustic, the change is $a_S \to 1 - iT/2$. On the other side $a_S \to 1 + iT/2$. As should be expected though, the precise assignation of a value to a_S in this region has little effect on the final outcome at the level of approximation used here.

A complete solution of the tunnelling problem is obtained by matching the solutions $\psi_\alpha(q)$ to form a globally single-valued function on λ. Equations (12)–(15) tell us how to compute the discrete amplitude changes that occur when we pass through a caustic to switch from one branch to another. For example, when passing from branch N to branch P, we must multiply the amplitude by the ratio

$$M_{N\to P} = \frac{a_P}{a_N} = e^{-i\pi/2}\,\frac{1+\frac{1}{2}iT}{1-\frac{1}{2}iT} \approx e^{-i\pi/2+iT}\,. \tag{16}$$

This represents a simple modification of the Maslov phase shift $e^{-i\pi/2}$. The tunnelling parameter T multiplies the exponentially growing solution on the forbidden side and must therefore be small—so the modification represents a small correction to the Maslov phase shift that vanishes when T vanishes. In the particular case where it is is possible to let q approach real infinity along the dominant branch, then we must choose $T = 0$, or else $\psi(q)$ fails to satisfy the boundary conditions. Then there is no correction to the Maslov shift. There can be cases, however, where the dominant branch is effectively bounded in real configuration space and then T can be non-zero.

One can similarly compute multipliers $M_{\alpha\to\beta}$ for the transition between any other pair of branches. In each case the multiplier depends on the variable T. Each fold caustic has associated with it its own T. To compute the T's and quantize the system, one considers every independent closed loop C in λ and demands that the total change in the amplitude, including a phase factor for the total action, reduces to unity as follows:

$$\exp\left[\frac{i}{\hbar}\oint_C p\cdot dq\right] \prod_{\substack{\text{transitions}\\ \alpha\to\beta\\ \text{along } C}} M_{\alpha\to\beta} = 1\,. \tag{17}$$

One gets one independent equation for each homologically independent loop and one tunnelling parameter for each fold caustic. Those tunnelling parameters that connect dominant solutions along which one can approach $q \to \infty$ can be set to zero, as discussed above, but the rest form a set of unknowns which are determined by solving the simultaneous set of equations obtained by applying (17) to every homologically independent loop. In addition one obtains any quantization conditions on the energy E by solving this set of equations.

We have concentrated here on the case of crossings of real caustics because the rules implied by (12)–(15) are easy to state and can be used to solve all the problems considered in this paper. The same ideas apply, however, when structure away from real q is important. Then multipliers come at crossings of Stokes surfaces—and the multipliers associated with

the three Stokes surfaces coming from a caustic will all depend on a tunnelling parameter that is fixed for each caustic. The rules are less easy to state for the general case, however, because there are no obvious phase conventions for the amplitudes of the solutions ψ_α (see the appendix).

It is instructive to see how the process described above compares with the example of standard real EBK quantization of tori in n-dimensional integrable systems. Here one gets n equations of the form (17) from the n independent loops C_a on the torus. Ignoring tunnelling effects, all the tunnelling parameters can be set to zero and the multipliers reduce to standard Maslov phase shifts. Equation (17) then reduces to the EBK quantization conditions

$$\frac{1}{2\pi}\oint_{C_a} p\cdot \mathrm{d}q = \left(n_a + \tfrac{1}{4}\mu_a\right)\hbar \qquad a = 1,\ldots,n. \tag{18}$$

The total Maslov index μ_a represents the sum of individual Maslov phase shifts from each caustic that C_a crosses. Like the product of multipliers in (17), it is not, at first sight, representation-independent. There is a deep result by Arnold [13], however, that shows that μ_a actually is independent of representation when computed over a closed loop—he shows that μ_a is a winding number in the space of Lagrangian planes. See also results specific to tori in Littlejohn and Robbins [11].

Likewise it should be possible to show that $\prod M_{\alpha\to\beta}$ is representation-independent for closed loops even when tunnelling parameters are included. Unfortunately, this has not been done explicitly—the non-local nature of the amplitude changes makes the calculation more difficult in the fully complex case. Every other aspect of the quantization—the topology of λ as expressed by its homologically independent loops and the actions of those loops—is representation-independent. It 'must' be true that the product of multipliers is also, and, besides making the calculation more satisfying from an aesthetic point of view, a fully representation-independent calculation will be more easily implemented in practice. For now, though, we are forced to compute $\prod M_{\alpha\to\beta}$ using the representation-dependent rules discussed above.

Now we will illustrate this procedure with a number of examples.

4. Example: double well in one dimension

We will first illustrate the notation and procedure on the canonical example of a one-dimensional double well oscillator. For the sake of concreteness we can take the potential to be the quartic,

$$V(q) = -\tfrac{1}{2}q^2 + \tfrac{1}{4}q^4 \tag{19}$$

and consider negative energies, so there is a splitting in energy levels due to tunnelling between the left and right wells.

The Lagrangian manifold determining the states here is the level set $p^2/2 + V(q) = E$, whose full structure as a complex manifold has been discussed in section 2 and illustrated in figure 1. Let us illustrate how the essential topology of the full complex manifold is captured by a slice through it corresponding to real q—this procedure is all that will be available to us in multidimensional calculations. The real slice is shown in figure 3. There are two real 1-tori corresponding to real dynamics within each of the two wells and corresponding to the loops C_1 and C_2 in figure 1(b). Connecting them at the inner turning points is a complex 1-torus, shown by a broken circle, and this corresponds to the loop C_T of figure 1(b). Also shown as broken curves are complex solutions running off to infinity from the outside turning points of C_1 and C_2.

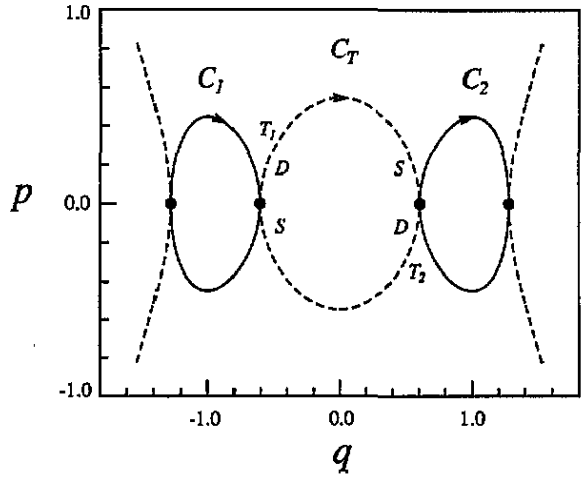

Figure 3. A real-q slice through the energy shell $E = 0.15$ for the potential in (19). Complex solutions are shown by broken curves and real solutions by full curves.

The tunnelling parameters associated with the outside caustics vanish according to the general discussion in the previous section. Let us associate tunnelling parameters T_1 and T_2 with the inner turning points of C_1 and C_2, respectively. Going once around each of C_1 and C_2 in the clockwise direction, one gets for (17),

$$\exp\left[\frac{\mathrm{i}}{\hbar}S\right]\mathrm{e}^{-\mathrm{i}\pi/2}\,\mathrm{e}^{-\mathrm{i}\pi/2}\frac{1+\frac{1}{2}\mathrm{i}T_a}{1-\frac{1}{2}\mathrm{i}T_a}=1 \qquad a=1,2 \tag{20}$$

where S is the common action of real loops C_1 and C_2. The factor $\exp[\mathrm{i}S/\hbar]$ is as in (17). The first factor of $\mathrm{e}^{-\mathrm{i}\pi/2}$ represents the Maslov shift for the outer caustic and the remaining terms represent the modified Maslov shift for the inner caustic. There is one remaining independent loop—around C_T—and (17) for this is

$$\exp\left[\frac{1}{\hbar}K\right]T_1\,T_2=1 \tag{21}$$

where $\mathrm{i}K$ is the action integral around loop C_T. K is a positive real number. Notice in figure 3 that if one leaves one turning point along a dominant branch, one approaches the other turning point along its subdominant branch. Then the multipliers along C_T are simply T_1 and T_2, as in (21).

Equations (20)–(21) represent three independent equations for the three unknowns, T_1, T_2 and E. The solution yields

$$T_1=T_2=T=\pm\exp\left[-\frac{K}{2\hbar}\right] \tag{22}$$

for the tunnelling parameters (there are two solutions corresponding to odd and even parity) and

$$\frac{1}{2\pi}S(E)=\left(n+\tfrac{1}{2}\right)\hbar-\frac{\hbar}{2\pi}T \tag{23}$$

then determines the corresponding two energy levels implicitly. To derive these equations one uses the fact that T_1, T_2, $\exp[-K/\hbar]$ are small. From here one gets the standard result for the splitting, $\Delta E=(2\hbar/\tau)\exp[-K/2\hbar]$, where τ is the period.

5. Example: double well in two dimensions

As a second illustrative example we examine the two-dimensional potential,

$$V(x, y) = -\tfrac{1}{2}x^2 + \tfrac{1}{4}x^4 + \tfrac{1}{2}y^2. \tag{24}$$

This is a separable system and can be solved by reduction to two one-dimensional systems and so is not interesting in its own right. However, this is the simplest possible system on which to demonstrate the general multidimensional calculation and the calculation contains all the essential features of the completely general case—the semiclassical discussion in this section will not make explicit use of the separability of the problem, just the fact that it is integrable.

This system has the two first integrals $H_x(x, p_x)$ and $H_y(y, p_y)$, obtained by extracting the x- and y-dependent parts of the Hamiltonian, respectively. Quantum states are obtained from the Lagrangian manifold defined by specifying the first integrals $H_x = E_x$ and $H_y = E_y$. Restricted to real phase space, this level set consists of two disjoint congruent tori corresponding to motion restricted to two symmetric wells on either side of the y-axis. EBK quantization of these tori yields doubly degenerate levels and this degeneracy is lifted when tunnelling effects are included. The tori are box tori—when projected to configuration space they fill out rectangles, as shown by the solid tori and shaded rectangles in figure 4.

The full complex manifold is four-dimensional and embedded in eight-dimensional phase space, so obviously a complete representation analogous to figure 1 is out of the question. We can, however, make a generalization of figure 3, and, as in the case of the one-dimensional quartic oscillator, this representation contains all the essential topological information. We solve for $p(q)$ with q real but p allowed to be complex. Except at caustics, there are everywhere four branches of $p(q)$. The four solutions are real in the shaded regions in figure 4 under the real tori. Everywhere else the solutions are complex. The interesting complex solutions are in the rectangle between the two real tori—there the solutions form a complex torus as shown by the dashed figure in figure 4. Along these branches one cannot approach $q \to \infty$ without passing through a caustic—so we can associate non-zero tunnelling parameters with caustics connecting onto this region. In all other

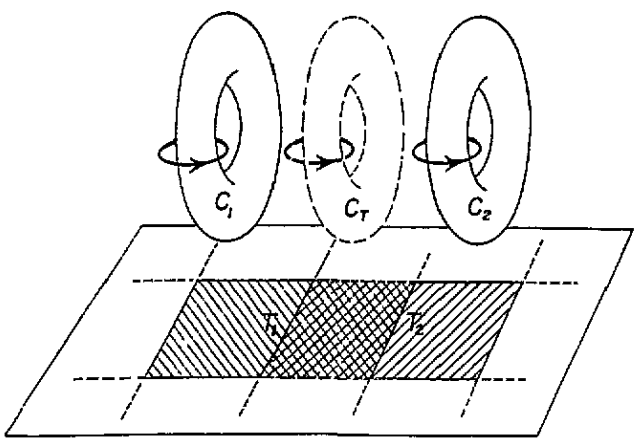

Figure 4. A real-q slice through the manifold for the potential in (24). Real structures are shown by full lines and complex by broken lines. The three interesting tori project to rectangles in phase space—two real and one complex one between them.

parts of configuration space it is possible to approach infinity along complex branches, so the relevant caustics have zero tunnelling parameters.

Let us now apply (17) to the topologically distinct loops. First we can encircle the tori along loops where (y, p_y) vary but (x, p_x) are held fixed. Here we just get the usual EBK quantization of the y-action because there are no non-trivial tunnelling parameters along these loops. Moving along the x-direction, however, we get the more interesting loops C_1, C_T and C_2 as shown in figure 4. The similarity between the labelling of these three loops and those of the previous section is not accidental—quantization of the loops follows exactly the calculation of the previous section for the one-dimensional double oscillator. Except here the picture going with the calculations is entirely two-dimensional. After following the same algebra, the x-action is quantized again according to (21)–(22).

At the end of this calculation we have simply reproduced the one-dimensional results for each separated Hamiltonian. But the important point is that we have not used the separability of the system anywhere. If a non-separable but integrable system had the same topology the calculation would be just the same. In the next section we will examine a non-separable but integrable system, albeit with a slightly different topology. The topology described in this section might also have relevence in more general problems, however. The essential feature is that there are two real box tori in symmetric wells connected by a complex torus in the middle—as if there were three tyres stacked together. This is the simplest possible way to connect two symmetric box tori and the structure may exist in other applications.

Another case is where the system above is perturbed to form a KAM system. Certainly the real tori persist under perturbation and it is well known that application of EBK quantization to such KAM tori works well. If the complex torus inbetween also persists then the calculation would also apply to this system—and here reduction to one-dimensional systems is not an option. This question will be the subject of a future publication.

6. Example: integrable but non-separable double well

To illustrate the quantization of general integrable systems, we calculate in this section splittings in an integrable, but non-separable system. The procedure for general integrable systems is as follows. One starts with some discrete number of congruent real tori, which are approximately quantized by the EBK quantization conditions (18). One then uses the first integrals to compute all branches of $p(q)$ over real configuration space, including complex branches, thereby connecting the distinct real tori across the forbidden regions. Some of the complex branches will run to ∞, and caustics connecting to these branches are assigned zero tunnelling parameters. Some other complex branches will form compact structures such as the tunnelling torus in figure 4—each distinct caustic connecting to these branches is assigned an initially unknown tunnelling parameter. The tunnelling parameters and splittings are then determined by application of 17 to all independent closed loops. This procedure only works if the important closed loops of the complex manifold are revealed by a real-q slice. More generally, however, it should be possible to deal with complex tunnelling routes by projecting to a different plane, such as momentum space. This is discussed further at the end of the section. Let us now look at the promised example.

6.1. The system

We consider the potential [15–17]

$$V(x, y) = -8x^2 + 8x^4 - 2y^2 + y^4 + 6x^2y^2 . \tag{25}$$

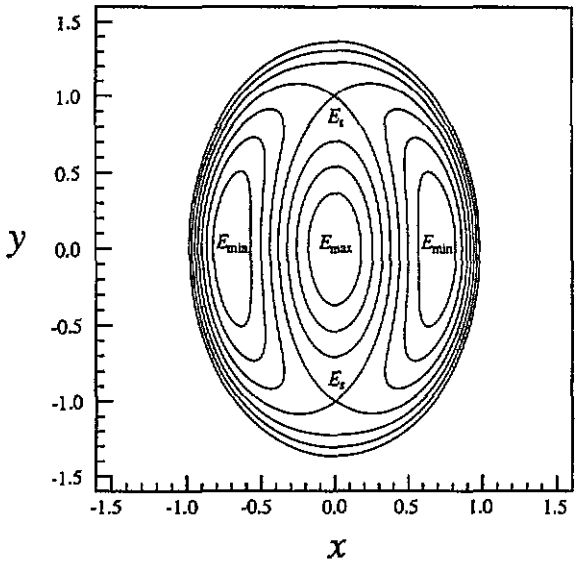

Figure 5. Contour plot of the potential in (25). There are two wells on either side of the y-axis. Also, there are two saddle points on the y-axis and a local maximum at the origin.

This potential is example 3.6 of reference [16]—we have interchanged the x- and y-coordinates and chosen the values $\kappa = -2$ and $\epsilon = \mu = \nu = \lambda = 0$ for the parameters appearing there. A contour plot of the potential is shown in figure 5. There are two symmetric wells on either side of the y-axis—as in the previous section there will be congruent tori localized within each of these wells and tunnelling will split the degeneracy in EBK levels. The potential has local minima of $E_{\min} = -2$ at $(x, y) = (\pm 2^{-1/2}, 0)$, saddle points of energy $E_s = -1$ at $(x, y) = (0, \pm 1)$ and a local maximum of energy $E_{\max} = 0$ at $(x, y) = (0, 0)$.

The quantum Hamiltonian has the following first integral [16]:

$$I = p_y^4 + 2(y^4 p_y^2 + p_y^2 y^4) + 12x^2(y^2 p_y^2 + p_y^2 y^2) - 4(y^2 p_y^2 + p_y^2 y^2) \\ -4(xp_x + p_x x)(y^3 p_y + p_y y^3) + y^2 p_x^2 + U(x, y) + 12\hbar^2 x^2 \qquad (26)$$

where the $O(\hbar^2)$ term is specific to the ordering convention and

$$U(x, y) = 4y^8 - 16y^6 + 16y^4 + 16x^2 y^6 - 32x^2 y^4 + 16x^4 y^4 . \qquad (27)$$

That is, I is a Hermitian operator that commutes with the Hamiltonian and is functionally independent of it in the classical limit. The integral is quartic in the momentum—separation of coordinates would lead to a first integral quadratic in momentum, so this system is not separable.

Let us restrict ourselves to negative energies, so that the centre of the potential is always energetically forbidden. Real tori exist when I is between 0 and $4(E - E_{\min})^2$. As I approaches 0 a torus shrinks to a stable periodic orbit that runs along the x-axis through one of the minima of the potential. At the other extreme, when I approaches $4(E - E_{\min})^2$ from below, a torus shrinks to a stable periodic orbit that intersects the x-axis vertically—moving parallel to the y-axis as it passes through a minimum of the potential. When $E < E_s$ the tori are energetically localized within each well irrespective of I, as with the tori in figure 6(a). For larger energies they are dynamically localized within a well when $I < 4E^2$ but encircle the local maximum completely when $I > 4E^2$ (as with the torus in figure 6(b)). In the case

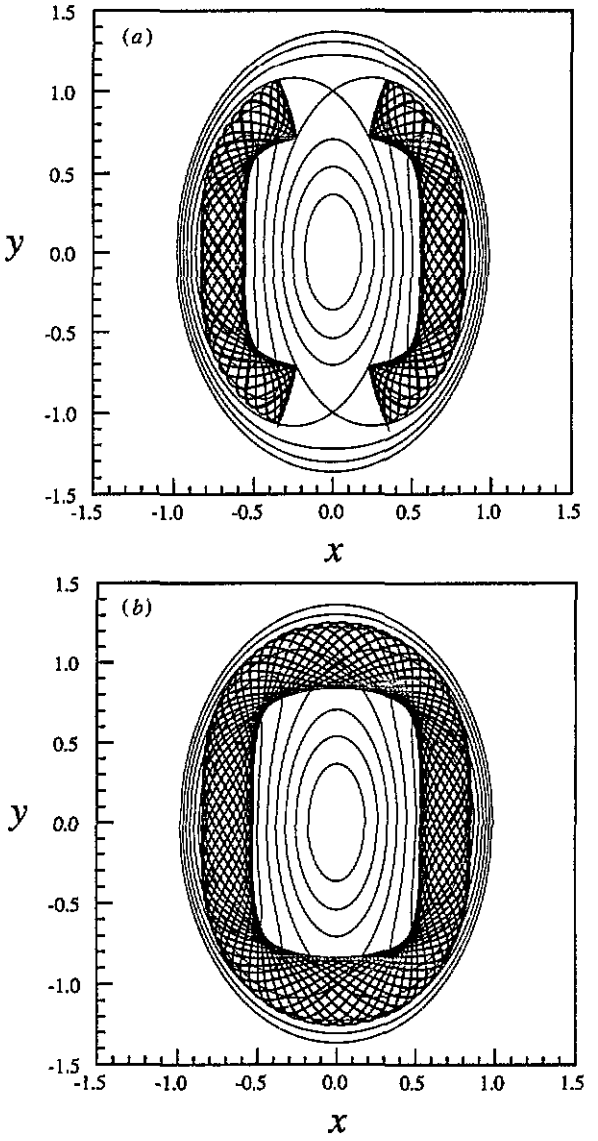

Figure 6. In figure 6(*a*) we show trajectories in configuration space on two spatially localized tori, for $E = -1$ and $I = 2$, related to each other by reflection in x. In figure 6(*b*) we show a torus for $E = -\frac{1}{2}$ and $I = 5$ that is not spatially localized. This one has a symmetric partner related to it through time reversal. Figure 6(*a*) is representative of the case $I < 4E^2$, and figure 6(*b*) of $I > 4E^2$.

$I > 4E^2$ there are still two degenerate tori but this time the symmetry relating them is time reversal not reflection in x—the tori encircle the local maximum in either the clockwise or anticlockwise directions—and each of these two tori shrinks to a periodic orbit encircling the centre as I approaches $4(E - E_{\min})^2$. We will deal with the case of spatially localized tori here but the same techniques could be applied to the case $I > 4E^2$ by projecting onto momentum coordinates instead of configuration coordinates.

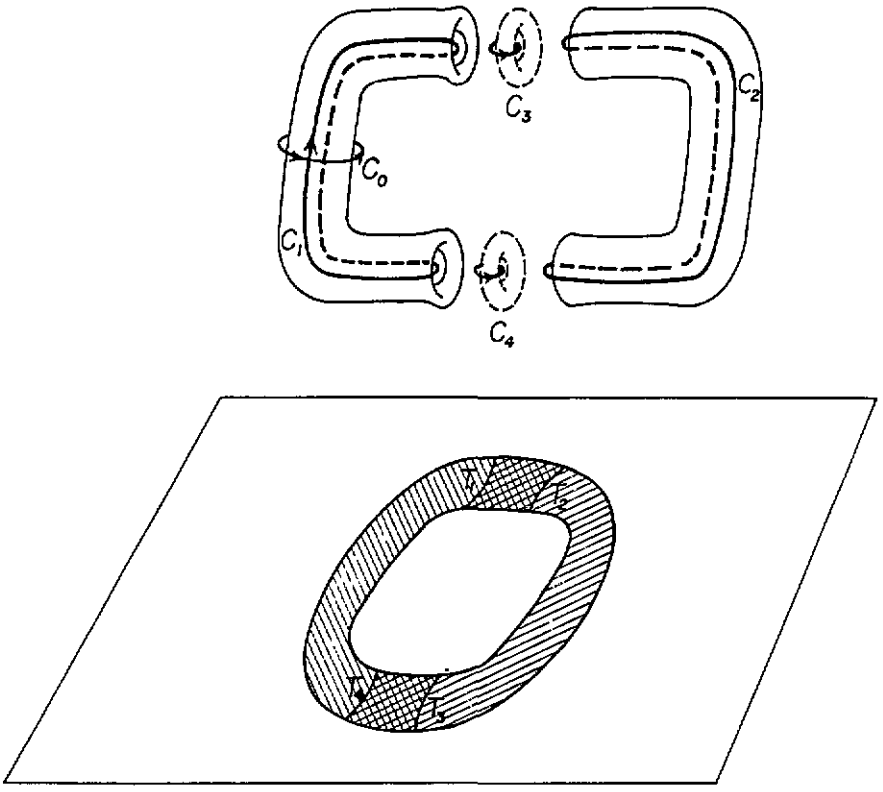

Figure 7. This is a schematic representation of the complex structure going with a torus of the form of figure 6(*a*). We only show the most important tunnelling tori. There are also tori connecting across the region around the x-axis but these are not as important for tunnelling. This figure is meant to be the analogue of figure 4.

In figure 6(*a*) we show a typical pair of spatially localized tori—in this particular case for $E = -1$ and $I = 2$. These two distinct real tori are both contained within one complex manifold and we will now pick out the features of the manifold that are relevant for tunnelling between them. Everywhere except at caustics there are eight branches of the complex manifold over configuration space, however only four of those turn out to be important at any given time. We illustrate the parts of the manifold that are important for the end result in figure 7, where we show schematically the two real tori, and two complex tori connecting the arms of the real tori. This simple structure, not much different from that of the previous section, explains the splittings between quantum states over much of this range of (E, I) values. Before describing more complicated tunnelling routes we will calculate splittings using just this part of the complex manifold.

The tori in figure 7 have two actions J_l and J_s, which we will refer to as the long and short actions, respectively. The long action J_l corresponds to any one of the loops C_1 and C_2 in figure 7. The short action J_s is defined by the loop C_0 on the left torus or its symmetric partner on the right. We will, for now, ignore tunnelling in applying (17) to the loop defining the short action and the result is the the EBK quantization condition,

$$J_s = \left(n_s + \tfrac{1}{2}\right)\hbar\,. \tag{28}$$

The Maslov index for the short loop is $\mu_s = 2$.

6.2. The quantization

There are five independent loops with which to quantize the long action. Four of the loops are obvious—left and right real loops C_1 and C_2, and top and bottom tunnelling loops C_3 and C_4. We introduce four tunnelling parameters, also shown in figure 7—T_1 and T_2 connecting the left and right tori through the top tunnelling torus, and T_3 and T_4 through the bottom tunnelling torus. Quantization of C_1 and C_2 yields two equations of the form

$$\exp\left[\frac{\mathrm{i}}{\hbar}S_l\right] \mathrm{e}^{-\mathrm{i}\pi/2}\frac{1+\frac{1}{2}\mathrm{i}T_a}{1-\frac{1}{2}\mathrm{i}T_a}\,\mathrm{e}^{-\mathrm{i}\pi/2}\frac{1+\frac{1}{2}\mathrm{i}T_b}{1-\frac{1}{2}\mathrm{i}T_b} = 1 \tag{29}$$

where $a, b = 1, 4$ for C_1 and $a, b = 2, 3$ for C_2. Quantization of C_3 and C_4 gives two equations

$$\exp\left[\frac{1}{\hbar}K_l\right] T_c T_d = 1 \tag{30}$$

where $c, d = 1, 2$ for C_3 and $a, b = 3, 4$ for C_4. As in previous sections, $\mathrm{i}K_l$ is the (common) action integral around loops C_3 and C_4, and K_l is a positive real number. There is a fifth homologically independent loop C_5. This loop goes along one branch of the left torus, transitions to the right torus through the upper tunnelling torus along the dominant branch, passes around the right torus, and back to the left torus along the dominant branch of the bottom tunnelling torus. The loops C_1–C_5 are complete in the sense that any closed loop that can be drawn around the long direction in figure 7 can be expressed as a linear combination of them. Quantization of C_5 yields

$$\exp\left[\frac{\mathrm{i}}{\hbar}S_l + \frac{1}{\hbar}K_l\right]\frac{T_1}{\mathrm{e}^{\mathrm{i}\pi/4}\left(1-\frac{1}{2}\mathrm{i}T_1\right)}\,\mathrm{e}^{-\mathrm{i}\pi/4}\left(1+\tfrac{1}{2}\mathrm{i}T_2\right)\frac{T_3}{\mathrm{e}^{\mathrm{i}\pi/4}\left(1-\frac{1}{2}\mathrm{i}T_3\right)}\mathrm{e}^{-\mathrm{i}\pi/4}\left(1+\tfrac{1}{2}\mathrm{i}T_4\right) = 1\,. \tag{31}$$

The multipliers here are each of the form $a_{\mathrm{D}}/a_{\mathrm{N}}$ or $a_{\mathrm{P}}/a_{\mathrm{S}}$ in the notation of (12)–(15). Equations (29)–(31) represent five independent equations for the five unknowns T_1, T_2, T_3, T_4 and E.

6.3. The solution

Using the fact that the tunnelling parameters are small, equation (29) immediately reduces to

$$J_l(E) = \left(n_l + \tfrac{1}{2}\right)\hbar - \frac{\hbar}{2\pi}T \tag{32}$$

where

$$T = T_1 + T_4 = T_2 + T_4\,. \tag{33}$$

We can use the rest of the quantization equations to solve for the tunnelling parameters, giving

$$T_1 = T_2 = T_3 = T_4 = \pm\exp\left[-\frac{K_l}{2\hbar}\right]. \tag{34}$$

So the splitting in action is $\Delta J_l = (2\hbar/\pi)\exp[-K_l/2\hbar]$ and the semiclassical prediction for the energy splitting is

$$\Delta E_{\mathrm{sc}}^{(l)} = \frac{\partial H}{\partial J_l}\,\Delta J_l = \frac{2\hbar\omega_l}{\pi}\exp\left[-\frac{K_l}{2\hbar}\right]. \tag{35}$$

Table 1. Comparison of the exact quantum splitting ΔE_{qm} with the semiclassical splittings $\Delta E_{sc}^{(l)}$ and $\Delta E_{sc}^{(s)}$ for some quantized tori at $\hbar = 0.06$. $\Delta E_{sc}^{(l)}$ is the splitting obtained from quantization of the long action and provides a good approximation to the quantum splitting—the splitting $\Delta E_{sc}^{(s)}$ obtained from quantization of the short action always represents a negligible correction.

n_l	n_s	E	I	$\Delta E_{sc}^{(s)}$	$\Delta E_{sc}^{(l)}$	ΔE_{qm}
2	1	−1.3086	0.1735	2.8×10^{-12}	1.033×10^{-8}	1.020×10^{-8}
4	1	−1.1601	0.4936	5.3×10^{-13}	7.187×10^{-5}	7.197×10^{-5}
6	1	−1.0206	0.9555	3.0×10^{-11}	3.342×10^{-5}	3.360×10^{-5}
2	2	−1.0025	0.1584	4.3×10^{-10}	7.755×10^{-7}	7.721×10^{-7}
4	2	−0.8630	0.4439	4.2×10^{-9}	3.343×10^{-4}	3.445×10^{-4}
6	2	−0.7350	0.8430	4.4×10^{-9}	9.922×10^{-4}	9.942×10^{-4}
2	3	−0.7126	0.1415	3.7×10^{-7}	3.508×10^{-5}	3.709×10^{-5}
4	3	−0.5846	0.3859	3.8×10^{-7}	1.017×10^{-3}	1.019×10^{-3}

In table 1 we compare this prediction with the exact splittings E_{qm} obtained from a numerical solution of the quantum mechanical problem for some select states. The actions and frequencies were computed for tori with E and I given by the exact quantum eigenvalues. The agreement is good, even though the quantum numbers chosen are quite modest, so the analysis offered here works.

It is not difficult to include tunnelling corrections in the quantization of the short action too, but the relevent complex action (K_s say) happens to be always larger in magnitude than K_l and makes little difference to the splitting. For the sake of completeness we give the extra correction anyway. If one cuts through the tori along a strip surrounding the x-axis, the result is structurally the same as that of an analogous strip in the potential of section 5. Quantization gives the same result found there, namely $\Delta J_s = (\hbar/\pi)\exp[-K_s/2\hbar]$ and a corresponding contribution to the energy splitting,

$$\Delta E_{sc}^{(s)} = \frac{\partial H}{\partial J_s}\,\Delta J_s = \frac{\hbar\omega_s}{\pi}\exp\left[-\frac{K_s}{2\hbar}\right]. \tag{36}$$

The total splitting is $\Delta E_{sc} = \Delta E_{sc}^{(l)} + \Delta E_{sc}^{(s)}$, but, as seen in table 1, $\Delta E_{sc}^{(s)}$ always makes a negligable contribution to ΔE_{sc}.

6.4. *General discussion*

The results discussed here indicate that tunnelling properties of integrable systems do not differ qualitatively between separable and non-separable systems. We have seen that similar calculations are used to predict the tunnelling splittings of this non-separable system and of the separable system examined in the previous section. In each case the discussion was based on the fact that the congruent tori are both contained within a single complex manifold, defined as a level set of the first integrals. Alternatively stated: when two congruent tori are analytically continued into the forbidden regions and to complex positions, they join together to form one smooth manifold. In references [5, 6] splittings were calculated under the assumption that the analytically continued tori join in the forbidden region in a different way. There it is also assumed that the continuations of the tori meet, but only along surfaces of one complex dimension (invariant under complex time dynamics).

A simple argument shows that generically the continued tori of exactly integrable systems either join to form a single smooth manifold or they do not meet at all. We give the argument for n dimensions. Then both of the tori and their continuations

are simultaneous level sets of the Hamiltonian H and $n-1$ additional first integrals $(A_2, \ldots, A_n)$. If a point p is contained in the intersection of the two manifolds, then the implicit function theorem guarantees that the continuations join to make a single smooth manifold unless the gradients $(\mathrm{d}H, \mathrm{d}A_2, \ldots, \mathrm{d}A_n)$ are linearly dependent at p. Linear dependence, $\mathrm{d}H + a_2\,\mathrm{d}A_2 + \cdots + a_n\,\mathrm{d}A_n = 0$, represents $2n$ conditions to be fulfilled, corresponding to the $2n$ dimensions of phase space, with only $2n-1$ free parameters to vary—the $n-1$ coefficients $(a_2 \ldots, a_n)$ plus another n for the n dimensions of the manifold. Therefore it is necessary to vary one (complex) parameter to find tori that intersect non-simply. The condition is fulfilled when the level set contains a separatrix.

One cannot easily use this argument for KAM tori in near-integrable systems, however. There, even though p is an analytic function of q for each individual torus, different tori are not related to each other analytically—the actions obtained from perturbation theory are not analytic functions of (q, p). Therefore one cannot easily construct globally analytic functions for which the congruent tori are simultaneous level sets, and the argument above does not hold. In [5] numerical evidence was presented which indicated that, in the case of near-degeneracy between two non-congruent tori in a KAM system, the continued tori do not join in a single manifold in this way. There, however, each torus has a completely independent structure, being perturbed by different sets of resonances in each case. In the case of symmetric tori the effects of perturbation are the same for both tori and it is conceivable that the continuations join simply, possibly in an approximate sense. Certainly, as an exactly integrable system is perturbed the tunnelling structure must initially be similar to that presented here. It has been found [18, 19] in a study of a class of quartic oscillators that the initial behaviour of the splitting between congruent tori in a perturbed system is consistent with a semiclassical explanation such as that posited here, though for larger perturbations coupling to chaotic states makes for a qualitatively different prediction for the splittings, with large fluctuations as a function of the perturbation parameter. We will investigate the structure of near-integrable systems in this context in a future publication.

The calculation presented here has relied on the fact that the closed loops which connect the real tori in the complex manifold can be detected by a slice through real configuration space. This is not always the case—as for tori of the type in figure 6(*b*) for the system examined in this section, where the congruent tori cover the same parts of configuration space and loops connecting them necessarily go to complex position. We have retained the advantage, however, that we do not rely on the Hamiltonian being of the kinetic-plus-potential type, and we are therefore free to project onto any Lagrangian plane (for example, momentum space) before embarking on the analysis. For example, for the torus in figure 6(*b*), projection onto either of the (p_x, y) or (p_x, p_y) planes reveals a clean tunnelling torus over real projection coordinates and the calculation can proceed as outlined in this and the previous sections.

Even this is not necessary in principle, however—if the loops connecting the real tori were not accessible in real coordinates for any projection one would just have to compute multipliers from a fully complex Stokes surface analysis—we have made this assumption merely to simplify the computation of multipliers and to help identify the relevant topology of the complex manifold. In fact, if and when a manifestly invariant interpretation for the product of multipliers over a closed loop is available, there will be no practical advantage in slicing through with real coordinates once the independent closed loops have been determined. (Of course, some sort of dimension-reducing procedure, such as taking a two-dimensional slice, would probably be necessary to identify the closed loops in the first place.)

7. Example: the case of H_2^+

The simplified model of the molecule H_2^+ where the single electron moves around two fixed protons is another integrable system with a double well. This time the Hamiltonian is separable in elliptic polar coordinates. The three-dimensional problem has already been solved in complete detail by Strand and Reinhardt [20] using separation of coordinates—including uniform approximations valid for tori that almost touch—so there is no point in going into detail here. However, because the system is such an important physical example of multidimensional integrable tunnelling, it is worthwhile explaining the intrinsically multidimensional topological structure governing this problem, and its connection with previous examples.

To keep things simple we restrict ourselves to two dimensions. The electron then moves in a two-dimensional potential,

$$V(x,y) = -\frac{e^2}{r_1} - \frac{e^2}{r_2} \tag{37}$$

where r_1 and r_2 are the distances from the electron to protons 1 and 2, which we suppose to be at positions $(x,y) = (-a,0)$ and $(x,y) = (+a,0)$, respectively. There exists the following first integral [3,20,21]:

$$I = \frac{L_1 L_2}{2a^2} + \frac{e^2}{a}(\cos\theta_1 - \cos\theta_2) \tag{38}$$

where L_1 and L_2 are the angular momenta of the electron about each of the protons, and θ_1 and θ_2 are the angles it makes with the x-axis at each of them. Real tori exist when I lies between 0 and $2e^2/a$. When, in addition, $I < -E$, there are two symmetric localized tori, shown schematically in figure 8.

An important part of the structure here is that the real tori have singularities in Cartesian coordinates at $r_1 = 0$ and $r_2 = 0$, and as a result it is difficult to give a fully three-dimensional schematic representation of them as we did for previous examples. We just show schematically the single tunnelling torus, and a projection onto configuration space of the real tori. Each real torus is bounded by an ellipse and a hyperbola, both with foci at the positions of the protons. We also show on the projection the paths of the loops corresponding to the actions $J_{\rm ell}$ and $J_{\rm hyp}$. The interesting action for tunnelling is $J_{\rm ell}$, for which the loop traces out in configuration space part of an ellipse with foci at the proton positions, as in figure 8. Both of the caustic crossings for this action connect the inner tunnelling torus. On the other hand, the loops for $J_{\rm hyp}$ trace out segments of a hyperbola and the only caustic crossings connect complex branches that run off to infinity, and for which the tunnelling parameters are therefore zero.

Quantization of the action $J_{\rm hyp}$ therefore just gives the EBK conditions without corrections. Quantization of $J_{\rm ell}$ in the left and right tori gives

$$\exp\left[\frac{\mathrm{i}}{\hbar}S_{\rm ell}\right]\left[\mathrm{e}^{-\mathrm{i}\pi/2}\frac{1+\frac{1}{2}\mathrm{i}T_a}{1-\frac{1}{2}\mathrm{i}T_a}\right]^2 = 1 \qquad a = 1,2 \tag{39}$$

where T_1 and T_2 are the left and right tunnelling parameters—both Maslov index shifts have here been modified in the same way through being connected to the same tunnelling torus. Solution for the T_a's follows the same route as for the J_l quantization in the previous section—the algebra is the same because all the tunnelling parameters are equal, though in the previous section they came from two distinct tunnelling tori and here only one. The result is the obvious equivalent of (35).

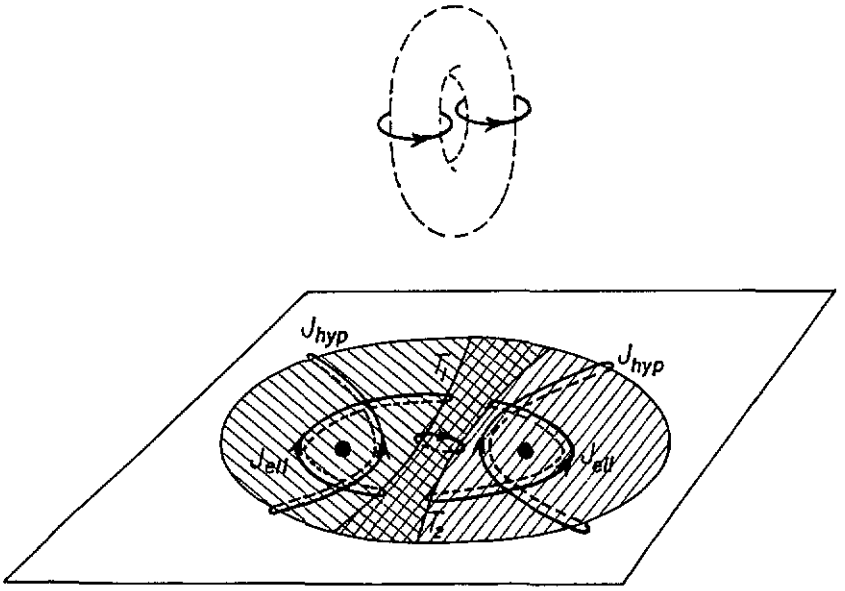

Figure 8. The real and tunnelling tori for two-dimensional $H_2{}^+$ in the regime where $I < -E$. The real tori are not drawn explicitly but they connect to the tunnelling tori, as the box tori do in figure 4. Here, however, the tori fall into the singularities to give the toral structure rather than having it be obvious from a simple projection.

As mentioned before, a complete analysis of the semiclassical quantization, including numerical comparisons, was performed in [20]. The description here shows how the solution falls into the general structure of tunnelling for integrable systems.

8. Conclusion

We have shown that tunnelling in integrable systems can be computed in an intrinsically multidimensional way, so that it does not rely on separation of coordinates. The solution comes from the application of a quantization condition to every homologically independent closed loop on the complex manifold that is defined by setting the first integrals to constants in complex phase space. In all of the examples presented here—in each case tunnelling was between two symmetric wells—the topological structure of the complex manifold could be analysed by a simple cut through it corresponding to real position, and the essential structure was that tori localized in each well were connected by one or more complex 'tunnelling tori' in the forbidden region. The essential structure was common to both separable and non-separable systems. Quantization of the loops on the real tori corresponded to modifications of the usual EBK quantization conditions, in which the Maslov indices were augmented by extra phase shifts that depended on unknown tunnelling parameters. Quantization of closed loops on the tunnelling tori determined the values of the tunnelling parameters and therefore completed the calculation of the corrections to the EBK conditions.

The calculation is almost canonically invariant, but not quite. The complex manifold defined by the set of first integrals is a coordinate-invariant object, as is the set of the homologically independent loops to which the quantization conditions are applied. The

quantization conditions (equation (17)) themselves, though, are not canonically invariant in the form discussed here. In equation (17) there appears an action integral over the closed loop, which is canonically invariant, and a product of multipliers that generalize the Maslov phase shifts at caustics, which is not. It remains an outstanding problem to demonstrate the invariance of the product of multipliers—this would represent a generalization of the proof by Arnold [4] that the sum of Maslov index shifts over a closed loop is an invariant. Invariance is more difficult to prove in the complex case because the multipliers are not determined by purely local calculations, as the Maslov indices are in the real case.

A further generalization of the multipliers might generate uniform approximations to splittings that are valid when the tunnelling torus has very small action. Such uniform approximations are well known in the one-dimensional [1] case and yield splittings that are valid through the top of a potential barrier. They have also been applied to the problem of H_2^+ by Strand and Reinhardt [20] through separation of coordinates. Here the appropriate generalization might be obtained by calculating a multiplier that connects both ends of a tunnelling torus at once rather than going through each caustic separately—such a calculation would yield a modified correction to the Maslov index. A similar modification of the Maslov index was used by Robbins *et al* [22] to calculate uniform approximations to the splittings in the rotational levels of the SF_6 molecule using periodic orbit theory.

Finally, we note that a potential application of these results is to the splittings between KAM tori in near-integrable systems. It is not obvious that the structure of real tori connected by tunnelling tori is preserved under perturbation, however, and numerical evidence by Wilkinson [6] indicates that for tunnelling between non-congruent tori the structure is not preserved. We will investigate this problem in a forthcoming publication.

Acknowledgments

I would like to thank Jarmo Hietarinta for suggesting the non-separable potential used in section 6, and Elizabeth Grothe for artistic input. I would also like to thank one of the referees for pointing out [7, 8]. This work was supported under the SCIENCE programme of the Commission of the European Communities under contract no S/SC1*/915161.

Appendix

Here we derive general rules for the computation of Stokes constants and use them to get the amplitudes shown in (12)–(15).

We start with one dimension, where two branches of the Riemann manifold coalesce simply at q_0 and p_0. A shift of coordinates will allow the turning point to be at $q_0 = p_0 = 0$—then, close to the turning point, the momentum of the two branches is of the form $p^2 = \beta q$, and the action is $S(q) = \frac{2}{3}\beta^{1/2}q^{3/2}$ when the integration is started at $q = 0$. We show the two Riemann sheets in figure A1, on which $p(q)$ and $S(q)$ are single-valued. The sheets are identified along the cuts labelled by j and k.

For a complete specification of the wavefunctions and computation of the Stokes constants, it is also necessary to adopt a phase convention for the amplitude. The amplitude is of the form

$$A(q) = f(q)\left[\frac{\partial p}{\partial q}\right]^{1/2} = g(q)\,q^{-1/4} \tag{A1}$$

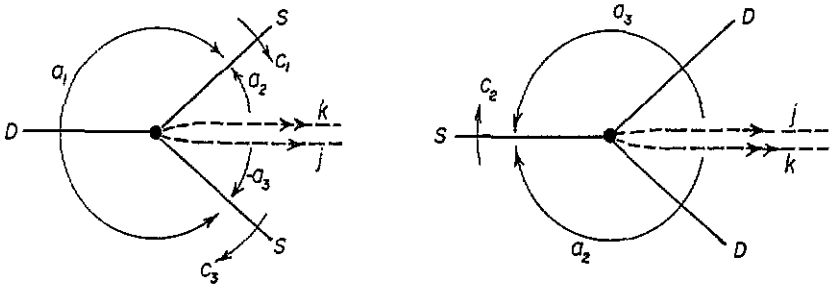

Figure A1. We show the Stokes lines on both of the Riemann sheets for a generic turning point—S for subdominant and D for subdominant. The sheets are identified along the edges labelled by j and k. The coefficents of the wavefunctions (a_1, a_2, a_3) depend on position as shown, changing at subdominant Stokes lines.

where $f(q)$ and $g(q)$ are analytic at the turning point. Because $A(q) \sim q^{-1/4}$, the amplitude is double-valued on the two sheets of figure A1, and it is necessary to insert a branch cut. Let us decide for the moment on the convention that when passing through the j side of the cut the amplitude changes sign—the sign does not change when passing through the other side of the cut, labelled by k.

The Stokes lines are shown by full lines in figure A1, and they are each labelled D and S according to whether they are on the dominant or subdominant branch. The coefficients of the solutions are denoted by a_1, a_2 and a_3 and these change discontinuously at subdominant Stokes lines. In figure A1 we indicate that the coefficient a_3 changes sign on passing through the j cut to account for the sign change in the fundamental solution. We denote the Stokes constants by c_1, c_2 and c_3 for clockwise crossings of the Stokes lines as in figure A1. The Stokes constants relate the different a's as follows:

$$a_2 = a_1 + c_1 a_3 \tag{A2}$$

$$a_3 = a_2 + c_2 a_1 \tag{A3}$$

$$a_1 = -a_3 + c_3 a_2 \,. \tag{A4}$$

These relationships must hold for arbitrary choice of any two of the a's and this implies that the Stoke constants are given by

$$(c_1, c_2, c_3) = (+1, -1, +1) \,. \tag{A5}$$

There is an asymmetry between the Stokes constants arising from the choice of branch cut for the amplitude on the double Riemann sheets.

This asymmetry can be eliminated by a different convention for the choice of phase of the amplitude. All that matters for the Stokes constant is the relative amplitude between the subdominant and dominant solution. Given a solution on the subdominant Stokes line, let us compute the Stokes constant relative to the dominant solution that is obtained by analytically continuing the subdominant solution once around the turning point in the clockwise direction. We also compute the Stokes constant with respect to clockwise crossing of the Stokes line. We use the same coefficients as in figure A1 to label the subdominant solution but the dominant solution may now suffer a sign change relative to the coefficients in the figure because of the altered phase convention. The Stokes constants then change the coefficients according to

$$a_2 = a_1 + c_1 a_3 \tag{A6}$$

$$a_3 = a_2 - c_2 a_1 \tag{A7}$$

$$a_1 = -a_3 + c_3 a_2 . \tag{A8}$$

The solution for this system is

$$(c_1, c_2, c_3) = (+1, +1, +1) . \tag{A9}$$

This is a much more transparent convention—the Stokes constants are all the same and they do not depend on a choice of branch cut for the amplitude.

The rules above can be very simply stated in a slightly generalized way as follows.

Rule. When passing through a Stokes line, determine the phase of the dominant solution by analytically continuing the subdominant solution around the singularity in the same direction in which the Stokes line is crossed. The Stokes constant is then $c = 1$.

For generalization to more dimensions there are two simple ways to apply these results. Firstly we can take the above arguments literally for the separated solutions $\chi_\alpha(q_1)$ of the normal coordinate decomposition (see equation (10)). Secondly, and perhaps more elegantly, we can just reinterpret figure A1 to represent (this time, schematically) the $2n$-dimensional complex manifold. The plane actually represents the $2n$-dimensional manifold, the lines $(2n-1)$-dimensional Stokes surfaces and the centre a $(2n-2)$-dimensional caustic surface. The figure also captures the essential structure in the multidimensional case and the calculations are then identical. In either case the multidimensional Stokes constants are given by the rule above, except now the word 'singularity' refers to a caustic surface instead of a turning point.

The discussion above works for a caustic in arbitrary position and orientation in complex configuration space. A very common case, however, is that the caustic intersects real configuration space and one is interested in calculating the wavefunction for real q. Then a more natural phase convention for the amplitudes is to choose them to be real and positive. We will now write down the connections between these different real amplitudes implied by the Stokes constants above. These formulae amount to a restatement of the Langer connection formula [2], but this time applied to Lagrangian manifolds in arbitrary dimensions.

First consider the solution that is purely subdominant on the forbidden side. It might lie, for example, on the Stokes surface labelled by the Stokes constant c_2 in figure A1. The branch cut would then lie on the allowed side. Let us continue the solution around the caustic to the allowed side. On continuing the solution in either direction, one encounters only Stokes surfaces on which it is dominant, so its own coefficient does not change—the other solution picks up a non-zero coefficient on the Stokes surfaces, however. So both solutions are present on the allowed side. Depending on the direction in which one passes around the caustic, the initially subdominant solution ends up either as the N branch or the P branch (labelled as in section 3). The solution that appears at the Stokes line gives the other. The coefficients of the N and P branches are obtained by analytically continuing the amplitude $A(q) \sim q^{-1/4}$ around the caustic in either direction and then re-expressing it in terms of the real positive amplitude $|A(q)|$. Expanding in normal coordinates one gets

$$|A_S| \exp\left[\frac{\mathrm{i}}{\hbar} S_S\right] \to \mathrm{e}^{\mathrm{i}\pi/4} |A_N| \exp\left[\frac{\mathrm{i}}{\hbar} S_N\right] + \mathrm{e}^{-\mathrm{i}\pi/4} |A_P| \exp\left[\frac{\mathrm{i}}{\hbar} S_P\right] . \tag{A10}$$

This is the same as for the one-dimensional discussion in Dingle [2].

Next consider a solution that is purely dominant on the forbidden side. Now, the forbidden side of the caustic at real q is actually on a Stokes surface of the caustic—so

one has to be careful about the coefficient of the subdominant solution since it undergoes a discontinuous change on passing through the Stokes surface. In a careful analysis of the change at a Stokes line [2] it is found that the coefficient on the Stokes line is the average of the coefficients on either side. In the conventions of figure A1 and (A2)–(A4) we have that $a_2 = -a_3$ and $a_1 = 2a_2 = -2a_3$. Re-expressing the solutions in terms of those with real positive amplitudes we get

$$|A_D| \exp\left[\frac{i}{\hbar} S_D\right] \to \tfrac{1}{2} e^{-i\pi/4} |A_N| \exp\left[\frac{i}{\hbar} S_N\right] + \tfrac{1}{2} e^{i\pi/4} |A_P| \exp\left[\frac{i}{\hbar} S_P\right]. \tag{A11}$$

Superposing (A10) with T times (A11) we recover the amplitudes in (12)–(15).

Notice that the connection formulae used in the calculation on real configuration space are not typical in that they relate solutions that lie exactly on Stokes surfaces. One could also traverse homological equivalents of the closed loops used in sections 4–7 that lay completely in complex phase space and only cross Stokes surfaces transverally—getting different but more generic multipliers at Stokes line crossings. The total product of the multipliers will be independent of the representative loop, however, and the final results will be the same.

References

[1] Berry M V and K E Mount 1972 *Rep. Prog. Phys.* **35** 389
[2] Dingle R B 1973 *Asymptotic Expansions: Their Derivation and Interpretation* (New York: Academic)
[3] Gutzwiller M C 1990 *Chaos in Classical and Quantum Mechanics* (Berlin: Springer)
[4] Arnold V I 1967 *Functional Anal. Appl.* **1** 1
[5] Wilkinson M 1986 *Physica* **21D** 341
[6] Wilkinson M 1987 *Physica* **27D** 201
[7] Meyer R E 1991 *SIAM J. Appl. Math.* **51** 1585, 1602
[8] Meyer R E and Shen M C 1992 *SIAM J. Appl. Math.* **52** 730
[9] Balian R, Parisi G and Voros A 1979 *Feynman Path Integrals (Lecture Notes in Physics 106)* ed S Alberverio (Berlin: Springer)
[10] Voros A 1983 *Ann. Inst. H. Poincaré* **39** 211
[11] Littlejohn R G and Robbins J M 1987 *Phys. Rev.* A **36** 2953
[12] Berk H L, Nevins W M and Roberts K V 1982 *J. Math. Phys.* **23** 988
[13] Arnold V I 1978 *Mathematical Methods of Classical Mechanics* (New York: Springer)
[14] Berry M V and Howls C J 1990 *Nonlinearity* **3** 281
[15] Grammaticos B and Dorizzi B 1984 *J. Math. Phys.* **25** 3470
[16] Hietarinta J 1984 *J. Math. Phys.* **25** 1833
[17] Hietarinta J 1987 *Phys. Rep.* **147** 87
[18] Bohigas O, Tomsovic S and Ullmo D 1993 *Phys. Rep.* **223** 45; 1990 *Phys. Rev. Lett.* **64** 1479; **65** 5
[19] Tomsovic S and Ullmo D *Preprint*
[20] Strand M P and Reinhardt W P 1979 *J. Chem. Phys.* **70** 3812
[21] Erikson H A and Hill E L 1949 *Phys. Rev.* **75** 29
[22] Robbins J M, Creagh S C and Littlejohn R G 1990 *Phys. Rev.* A **41** 6052